



生成 AI で教育を加速！最新トレンドと実践ガイド

Amazon Bedrock で Claude 3 を使いこなそう！

アマゾン ウェブ サービス ジャパン合同会社

パブリックセクター技術統括本部

シニアソリューションアーキテクト

松井 佑馬

自己紹介 松井 佑馬

パブリックセクター技術統括本部

教育・研究技術本部シニアソリューションアーキテクト



- 2014年 ソフトウェアエンジニアとして Amazon Japan 入社
その後プロダクトマネージャーにロール変更



- 2017年 Alexa チームに SA として異動し、様々なAlexa スキル開発を支援



- 2021年 AWS Japan に EdTech 担当 SA として入社
教育関係のお客様の技術支援を担当

アジェンダ

- 教育における生成 AI の活用
- Amazon Bedrock の概要
- Claude 3 の特徴と利用例

教育における生成 AI の活用



AWS は教育の DX を加速

テクノロジーとデータを民主化し、
子供たち一人ひとりの可能性を引き出す

教育のデジタル化



教育業界を取り巻く課題

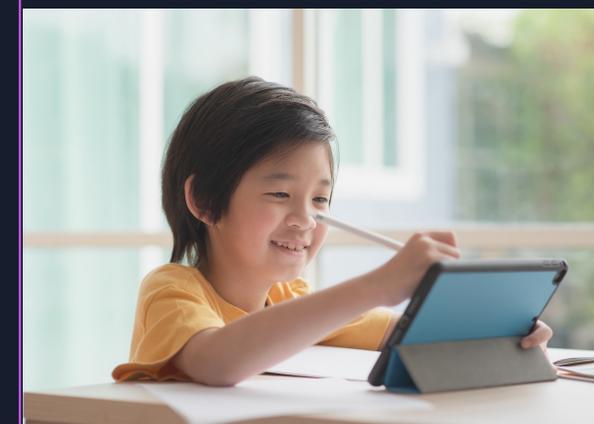
- **個別最適な学びと協働的な学びの実現**
 - デジタル技術を活用した効率的、効果的な学びの実現
 - データを活用した学習者一人ひとりに最適化された学びの実現
- **先生方の働き方改革の実現**
 - デジタル技術を活用した事務作業の負荷軽減
 - 学習者一人ひとりに向き合った創造的な学びの実践

生成 AI の活用で

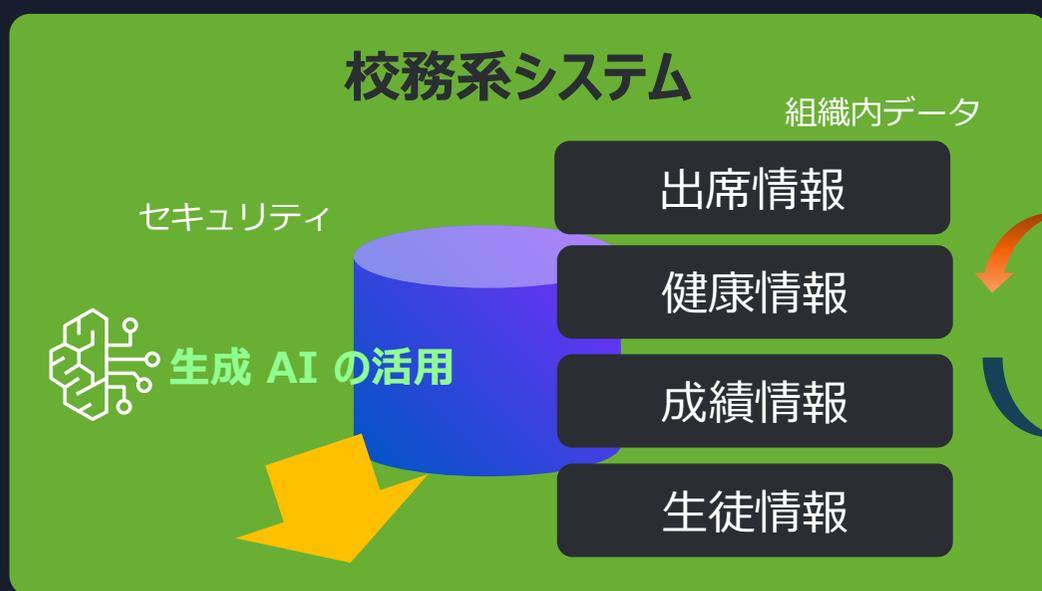
組織内のデータ x 生成 AI

教育における課題解決を加速

教育のデジタル化



AWS が支援したい・実現したい 教育 DX (公教育の例)



① 先生方の働き方改革の実現

② 個別最適な学びと協働的な学びの実現

組織内のデータ x 生成 AI による課題解決



組織内のデータ x 生成 AI で実現可能なユースケース例

 文書の要約

 記録からレポートの生成

 カリキュラム・時間割生成

 画像認識：写真の認識やタグ付け
書類の読み取り

校務

 テスト問題の生成

 テストの自動採点

 学習補助アシスタント

 画像生成：教材の素材画像
プレゼンテーションの素材画像

学習

テストの問題を自動で生成する

教員の月当たりの時間外労働時間は95時間にも上るとの調査もあり、社会問題と化している（参考）

このソリューションは、教員の労働状況改善に留まらず、生徒にとってより良い学習体験をもたらす可能性を秘めている

課題

より詳しくユースケースのアイデアや構成を知るには「公共機関における生成AIの活用案」を参照ください
 (QRコードをスキャンするか検索)

- 授業・部活・生活指導など、さまざまな業務をこなす中で、テストや課題の問題作成は大きな時間を費やす業務の一つである
- 問題集などの問題をそのまま利用すると、学習範囲や生徒の理解度に応じた問題を作るのは難しい上、その数にも限りがある

Point

- Amazon Kendraを用いて過去の問題や教材など参照し、新テスト問題を生成することができる
- 問題のポイントや解説を含めて生成できる
- 生成した問題を保存し、次の問題を作成する際の参照元にできる

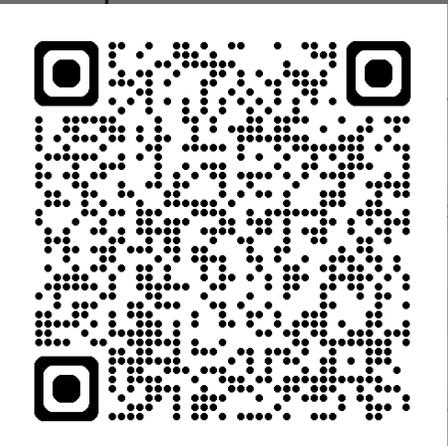
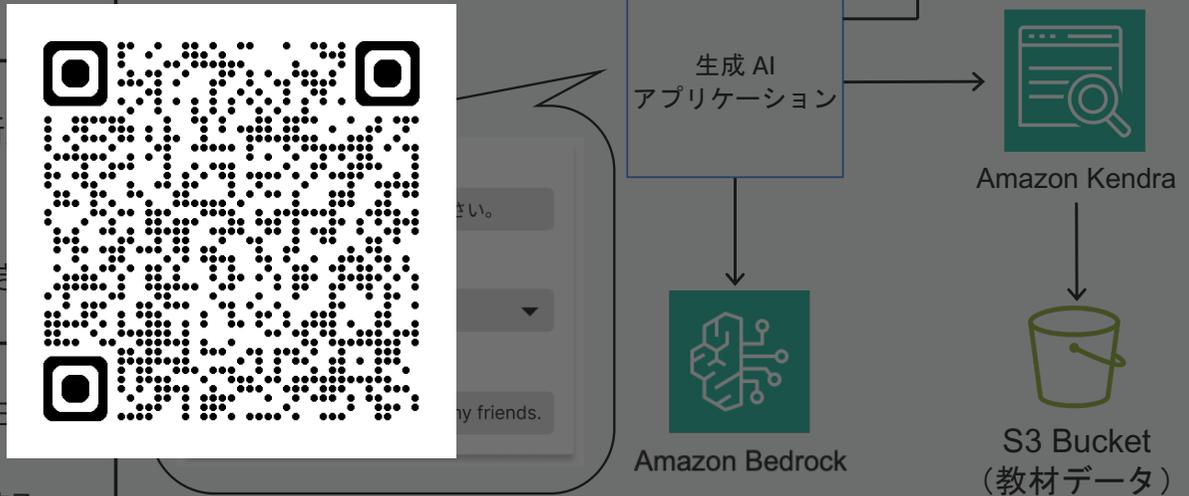
Next Step へのヒント

- 生成された問題を実際に解いた時の正答率やフィードバック・学習熟度を元に次の問題を生成する
- 生徒が自分自身で問題を作成して解けるようなアプリを作成する

ユースケース

中二範囲の不定詞の問題を作りたい

生成した問題の保存



Amazon Bedrock の概要



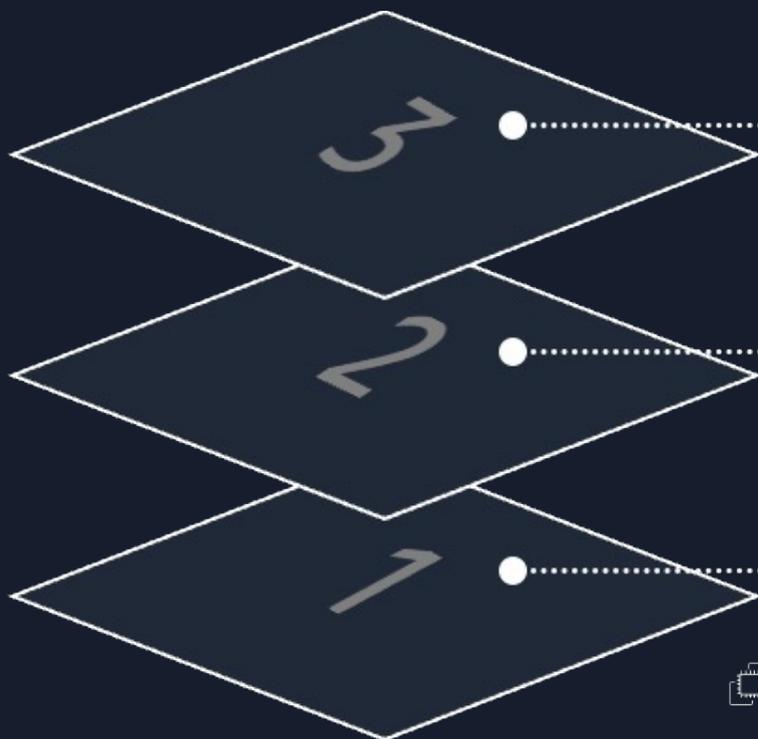
生成 AI とは？

- 会話、ストーリー、画像、動画、音楽など、新しいコンテンツやアイデアを創造する AI
- 一般に 基盤モデル (Foundation Model) と呼ばれる膨大なデータに基づいて事前にトレーニングされた大規模モデルを利用
- 基盤モデルは幅広いタスクへの適用が可能



業務に適した生成 AI ツールの選択

AWS の生成 AI 技術スタック



3. 基盤モデルを活用するアプリケーション

Amazon Q | Amazon Q in Amazon QuickSight | Amazon Q in Amazon Connect | Amazon CodeWhisperer

2. 基盤モデルを活用してアプリケーションを構築するツール群

Amazon Bedrock
ガードレール | エージェント | カスタマイズ機能

1. 基盤モデルの学習と推論のための基盤

GPUs | Tranium | Inferencia | SageMaker
Ultra Clusters | EFA | EC2 Capacity Blocks | Nitro | Neuron

シンプルで強力、
すぐに使える

細かいレベルの
制御と柔軟性



Amazon Bedrock

基盤モデルによる 生成 AI
アプリケーションを構築して
スケーリング

単一の API で主要な基盤モデルにアクセス

サーバーレス

セキュリティ・プライバシー・安全性

独自知識を活用した生成 (RAG)

複雑なタスクを実行可能なエージェント

Amazon Bedrock

様々な基盤モデルを用途に合わせて選択可能

AI21labs

amazon

ANTHROPIC

cohere

Meta

MISTRAL AI

stability.ai

テキスト生成用
多言語対応 LLM

テキスト要約、生成、
Q&A、検索、画像
生成

要約、複雑な推
論、執筆、コー
ディング

テキスト生成、
検索、分類

質疑応答と
読解力

テキスト要約、
Q&A、テキスト
分類、テキスト補
完、コード生成

高品質の画像と
アートの生成

Jurassic-2 Ultra

Amazon Titan Text Lite

Claude 3 Opus

Command

Llama 3 8B

New

Mistral Large

Stable Diffusion XL1.0

Jurassic-2 Mid

Amazon Titan Text Express

Claude 3 Sonnet

Command Light

Llama 3 70B

New

Mistral 7B

Stable Diffusion XL 0.8

Amazon Titan Text
Embeddings

Claude 3 Haiku

Embed English

Llama 2 13B

Mixtral 8x7B

Amazon Titan Text
Embeddings V2

New

Claude 2.1

Embed Multilingual

Llama 2 70B

Amazon Titan Multimodal
Embeddings

Claude Instant

Command R+

New

Command R

New

Amazon Titan Image
Generator

New

Amazon Bedrock データを安全かつ プライベートに



お客様のデータはいずれも、
基盤モデルのトレーニング
には使用されません



すべてのデータは暗号化され、
カスタマイズに使用されるデータは
お客様の VPC を介して安全に転送可能



GDPR、SOC、ISO、CSA
コンプライアンスおよびHIPAA
適格性サポート



Guardrails for Amazon Bedrock

アプリケーション要件と、責任ある AI ポリシーに合わせてカスタマイズされた保護機能を実現



Amazon Bedrock の複数の基盤モデルとエージェントにガードレールを適用



責任ある AI ポリシーに基づいた不適切なコンテンツのフィルタリング



短い自然言語の説明による拒否トピックの定義とその禁止



個人情報やカスタムで定義した正規表現による、機密情報の編集およびブロック

Claude 3 の特徴と利用例

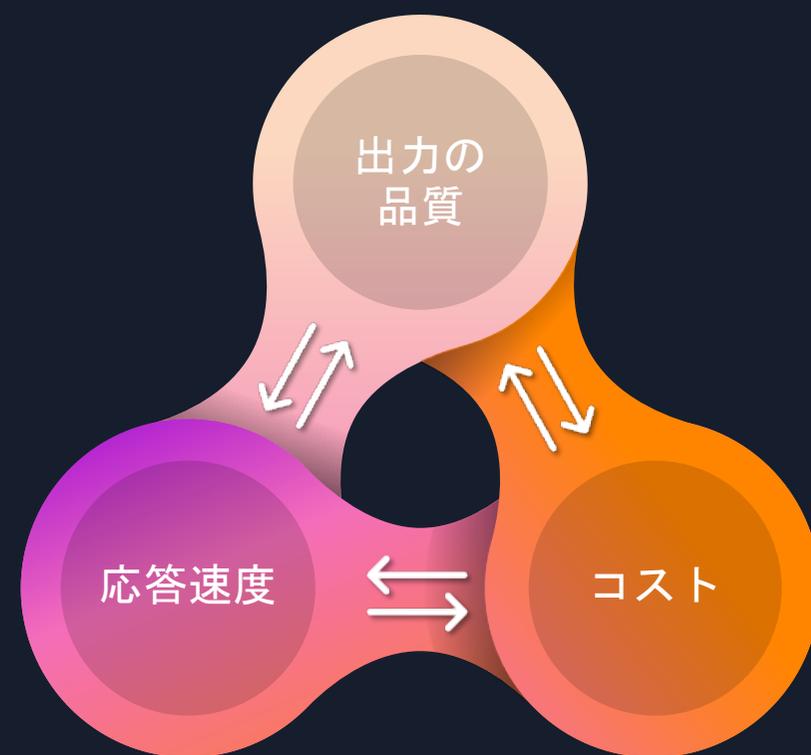


ニーズに合った最適なモデルを選択

Anthropic の最新基盤モデル Claude 3

	Claude 3 Opus	Claude 3 Sonnet	Claude 3 Haiku
主なユースケース	Anthropic の最も強力な AI モデルで、非常に複雑なタスクでもトップレベルのパフォーマンスを発揮	特にエンタープライズワークロードにとって、インテリジェンスとスピードの理想的なバランスを実現	ほぼ即時の応答性を実現する Anthropic の最速かつ最もコンパクトなモデル
Context	20万	20万	20万
画像	✓	✓	✓
コスト*	\$0.015 \$0.075	\$0.003 \$0.015	\$0.00025 \$0.00125

*Per 1K tokens



Claude 3 各モデルのユースケース

Claude 3 Opus

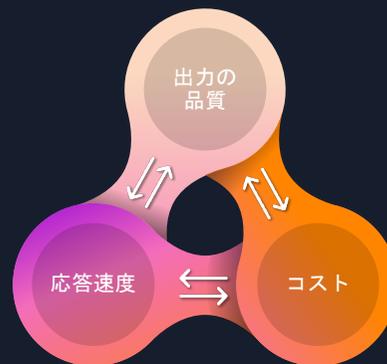
- 最も高性能なモデルが求められる場合
- 高度な論理的推論、複雑なコーディングやデータ分析、ツールを駆使する AI エージェント

Claude 3 Sonnet

- 応答速度やコストを含めバランスの良い高性能なモデルが求められる場合
- 高度なチャットボットや高品質な文章・画像の解析など、幅広いユースケースでの利用を想定

Claude 3 Haiku

- 応答速度の速さやコストの低さが求められる場合 (Haikuでも十分に対応可能なケースあり)
- チャットボット、データ抽出や分類、要約、画像分析、翻訳、RAG アプリケーション



Claude 3 の特徴 (1) Haikuは性能を維持しつつ高速に動作

書き起こし の要約

チャット

anthropic.claude-3-haiku-20240307-v1:0



▶ システムコンテキスト

アセは以上となります。山を振り替えていただきます。
デモでは、複数の写真から指定した人の写真をピックアップ
することができました。活用として以下のことができるかと思
います。教師写真のやま対応の支援卒業アルバム作成支援
大量の写真からの検索活用するためには、元となるデータが
重要であるため、顔検索の機能だけでなく、その他のサー
ビスや製品と組み合わせることが大切だと考えます。ファイ
ルサーバーとの連携もあるかと思います。ファイルサーバー
と連携を行うことで、今まで見落として写真を見つけること
ができるかもしれません。ACIサービスを活用した事業効率化
のデモセッションは以上となります。これらのサービスに興
味がありましたら、ご連絡を頂ければ幸いです。それではあ
りがとうございました。



Claude 3 の特徴 (2) マルチモーダル (テキスト+画像) 対応

- Claude 3 は 3 種類全てがマルチモーダルなモデル
- テキストのみで利用することも複数枚の画像を含めることも可能
- ユースケース
 - 資料や図、表からデータを抽出
 - 画像の状況分析
 - 手書きのメモや図表の文字起こし
 - 画像から製品情報や知見を生成



この車の状況を説明して



このレポートを要約して

Claude 3 の特徴 (2) マルチモーダル (テキスト+画像) 対応

画像のタグ付けと
説明文生成

チャット

anthropic.claude-3-haiku-20240307-v1:0



システムコンテキスト

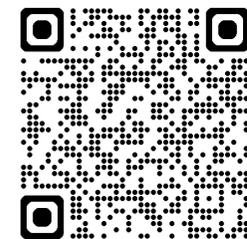


画像は修学旅行で撮った写真です。この写真に適切なタグを5つ出力して、また写真の分かりやすい説明文を付けてください。



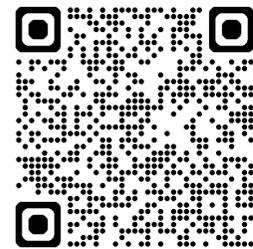
Claude プロンプトエンジニアリング

Anthropic
プロンプト
ライブラリ



1. **明確かつ直接的であること**: 応答を導くための明確な指示と背景を提供
2. **例を使用する**: プロンプトに例を含めて、目的の出力形式またはスタイルを指示
3. **役割を与える**: 特定の役割 (エキスパートの役割など) を割り当て
4. **XML タグを使用する**: タグを組み込んでプロンプトと応答を構造化し明確化
 - 「動物の名前を教えるので、その鳴き声で反応してください。<animal>XXX</animal>」
5. **段階的に考えさせる**: 出力の品質向上のために段階的な思考を奨励
 - 「・・・質問に答える前に、<thinking></thinking>タグの中で段階的に考えてください。その後、最終的な回答を<answer></answer>タグで出力してください。」
6. **出力形式の制御**: 一貫性と読みやすさのために目的の出力形式を指定
 - JSON/XML/HTML/CSV/Markdown/カスタムフォーマット等

Claude 3 を使いこなす: Generative AI Use Cases JP



デプロイした瞬間から
本番運用できる
日本語生成AIソリューション



ユーザー向けのチャットや要約・翻訳等
汎用的なユースケースをすぐに提供



高いカスタマイズ性



各ユーザーの使用履歴の全てを
アカウント内に閉じて自動で記録



認証や IP アドレス制限を
始めとしたセキュリティの担保



まとめ



まとめ

✓ 教育における生成 AI の活用

- 組織内のデータ x 生成 AI による学びや働き方の課題解決

✓ Amazon Bedrock の概要

- セキュアに基盤モデルを選択・活用してアプリケーション構築

✓ Claude 3 の特徴と利用例

- 高精度なものから高速なものまでマルチモーダル対応の基盤モデル
- プロンプトエンジニアリングおよびサンプルアプリケーション



SUMMIT JAPAN

日本最大の "AWS クラウドを学ぶイベント"

2024年6月20日、21日 | 幕張メッセ | 参加無料

- 初心者から経験者まで学べる！150以上のセッション
- ソリューション展示、業界別に特化した事例展示など250以上の展示ブース
- その場で導入や技術に関する相談をエキスパートにできるAWS導入&技術相談
- ライブステージ、展示、コミュニティをはじめとするDeveloper Zoneや、人気の体験型プログラムが勢揃い！



<https://aws.amazon.com/jp/summits/japan/>



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



公共機関における
生成 AI の活用案



Anthropic
プロンプトライブラリ



Generative AI
Use Cases JP



AWS Summit
Japan



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

(QRコードをスキャンするか検索して参考情報にアクセス)



生成 AI で教育を加速！最新トレンドと実践ガイド

生成 AI をお客様のアプリに組み込む

梅田 昌太

アマゾン ウェブ サービス ジャパン合同会社
技術統括本部 ソリューションアーキテクト

自己紹介

- 氏名: 梅田昌太
- 経歴: スタートアップや事業会社などでAWSを利用した開発に従事
- 現在は教育機関のお客様を技術支援
- 好きなサービス Amazon S3, AWS Glue



本日本話しする内容

- 生成AIをお客様アプリへ組み込むヒント
- RAGというアプローチ
- Knowledge bases for Amazon Bedrockのご紹介





Amazon Bedrock

基盤モデルによる 生成 AI
アプリケーションを構築して
スケーリングする最も簡単な方法

単一の API で主要な基盤モデルにアクセス

基盤モデルのカスタマイズ

検索拡張生成 (RAG)

複雑なタスクを実行可能なエージェント

セキュリティ・プライバシー・安全性



Amazon Bedrock

基盤モデルによる 生成 AI
アプリケーションを構築して
スケーリングする最も簡単な方法

単一の API で主要な基盤モデルにアクセス

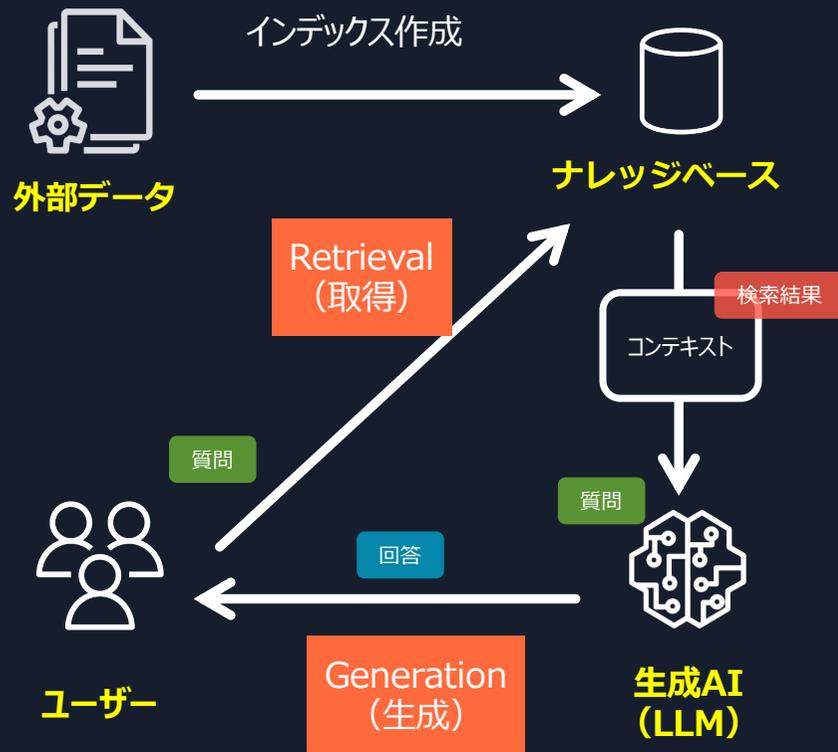
基盤モデルのカスタマイズ

検索拡張生成 (RAG)

複雑なタスクを実行可能なエージェント

セキュリティ・プライバシー・安全性

RAG (Retrieved Augmented Generation) とは



Retrieve: 外部データを元に作成されたナレッジベースから、質問（入力クエリ）に関連する情報を検索・取得する機能。

Generation: 検索された関連情報と入力クエリを組み合わせ、LLMを用いてそれらに基づく出力を生成する機能

RAG は Retrieve（検索）と Generation（生成）の2つのステージを組み合わせることで、幅広い知識を有効活用しながら、**高品質な回答を生成**するアプローチ

基盤モデルをカスタマイズする方法

プロンプト
エンジニアリング

検索拡張生成
(RAG)

ファイン
チューニング

継続事前学習

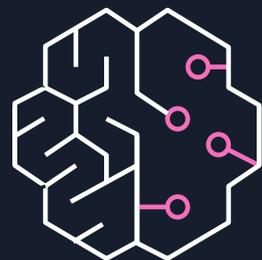
↑ 複雑性
質
価格
時間

RAG - 背景：「幻覚 (hallucination)」問題

大規模言語モデル(LLM)の特性として“もっともらしい回答”を生成してしまう
→ ビジネスでの利用を阻む壁 になりがち

Q: 佐藤先生の担当教科は何ですか？

A: ???



大規模言語モデル
(LLM)



“幻覚”が発生する主な理由：

- 学習データの偏り・不足
- 確率的な出力を行うモデルの特性
 - (一定の調整は可能だが)
モデルの出力は確率的なもの
- 評価指標の問題
 - 出力が正確かどうかは
そのLLM自体は評価しない

ビジネス利用においては、「幻覚」を含んだ出力を許容できない場合が多い

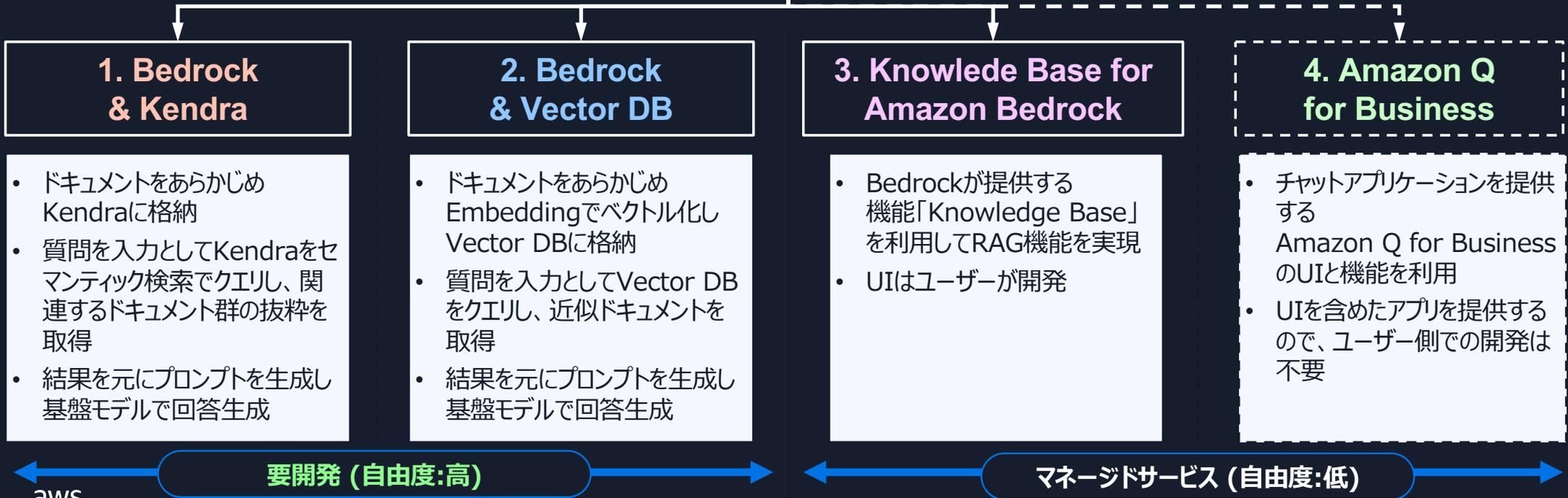
RAGのメリット

- **Hallucination抑止**：ナレッジベースから取得する事実から回答を生成することで、Hallucination（幻覚）をできるだけ抑止する
- **回答精度の向上**：外部情報を取り入れることで、モデル自体の限界や知識の欠如を補い回答の精度を向上できる（質問に関連性の高い回答を生成可能）
- **最新情報の反映**：ナレッジベースのデータを更新することで、最新の情報を反映させることが可能
- **出力の一貫性**：ナレッジベースからの検索結果をベースに回答することで、質問に対して一貫性のある回答を行うことが可能
- **応用範囲の広さ**：様々なタイプの問いに対して関連する情報を提供することで、幅広いアプリケーションでの応用が可能
- **LLMのカスタマイズ**：FineTuning を行うことなく、LLMが生成する回答に業界や組織固有の知識を判定することができる

RAG - AWSサービスを活用した実装アプローチ

RAG(検索拡張生成)のナレッジソース実装にはいくつかの選択肢がある

RAG (Retrieval Augmented Generation) の実装アプローチ



Knowledge bases for Amazon Bedrock

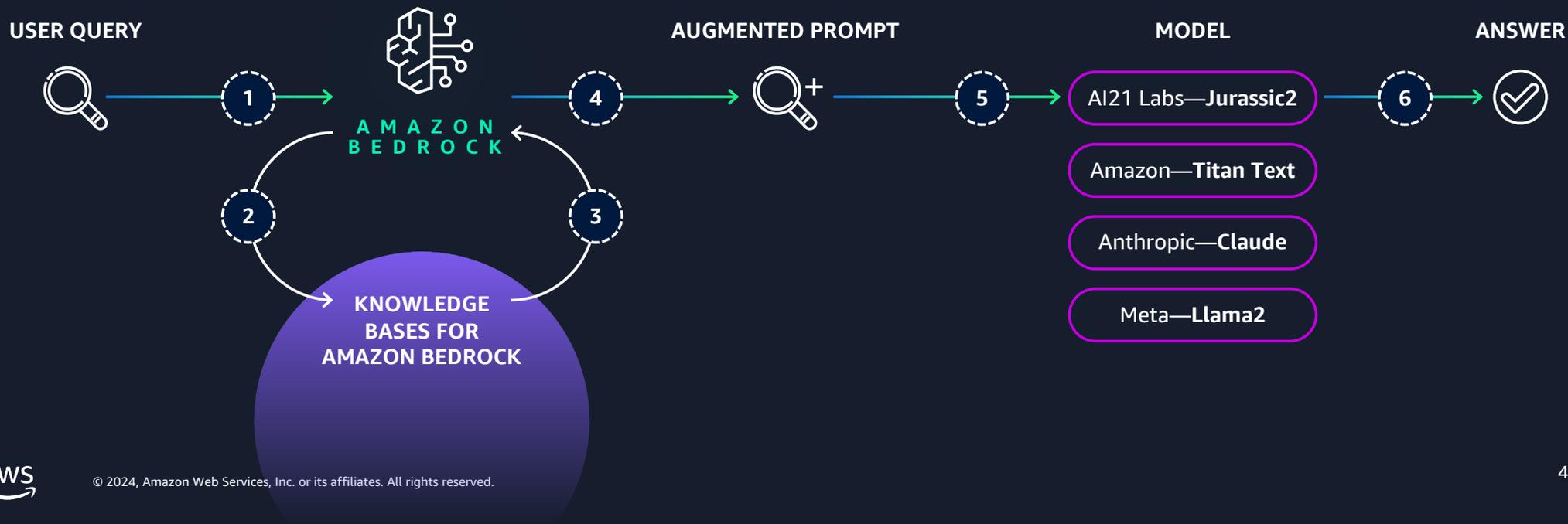
検索拡張生成 (RAG) のネイティブサポート

FMをデータソースに安全に接続し、RAGでより適切な回答を提供

取り込み、検索、生成を含むフルマネージド RAG ワークフロー

複数ターンにわたる会話のための組み込みセッションコンテキスト管理

検索機能付きの自動引用による透明性の向上



Knowledge bases for Amazon Bedrock

EXAMPLE : 教科担当の先生を教えてください

教師名	担当科目
佐藤先生	数学
	英語
鈴木先生	国語
	英語
高橋先生	理科
	社会
田中先生	英語

Knowledge bases for Amazon Bedrock



教師名

担当科目

佐藤先生

数学

英語

鈴木先生

国語

英語

高橋先生

理科

社会

田中先生

英語

Knowledge bases for Amazon Bedrock

組み込みSample

```
client = boto3.client('bedrock-agent-runtime', region_name="us-east-1")
response = client.retrieve_and_generate(
    input={
        'text': '佐藤先生の担当教科を教えてください'
    },
    retrieveAndGenerateConfiguration={
        "knowledgeBaseConfiguration": {
            "knowledgeBaseId": "TESTID",
            "modelArn": "arn:aws:bedrock:us-east-1::foundation-model/anthropic.claude-3-sonnet-20240229-v1:0"
        },
        "type": 'KNOWLEDGE_BASE',
    }
)
teacher = response["output"]["text"]
```

240 を超えるクラウドサービスであらゆるワークロードをサポート



コンピューティング



モバイル



ARとVR



エンドユーザーコンピューティング



ストレージ



データベース



ネットワークとコンテンツ配信



AWS コスト管理



機械学習



IoT



ロボット工学



ビジネスアプリケーション



メディアサービス



分析



マネジメントとガバナンス



開発者用ツール



サーバーレス



アプリケーション統合



Game Tech



量子テクノロジー



カスタマイズ可能なソリューション



移行と転送



ブロックチェーン



セキュリティ・ID・コンプライアンス



人工衛星



コンテナ





Thank you!



Classmethod Showcase

生成AIの信頼性を上げてビジネス を加速させるためのデータ活用術

新規事業部
熊谷 敏宏

自己紹介



熊谷敏宏 (Kuma**GAI** Toshihiro)

新規事業部

生成AIチーム

ソリューションセールス

経歴

- 2016年クラスメソッド入社、在籍8年目
- 最新のIT技術に目がない技術大好きセールス
- クラウド、アプリ開発、SaaSを経て、生成AI領域の専任営業を担当

趣味

- 趣味はランニング、マラソン、トレイルランニング (フルマラソン記録 3:26:00)
- 実家が飲食をされていて料理が得意

クラスメソッドの生成AI活用支援サービス

生成AIの専門家による技術支援からビジネス課題を解決するためのPoC支援、教育支援など幅広い支援を行っています。

支援領域	ご提供内容
生成AI環境構築サービス	社内で生成AIを活用するための独自環境構築を行います。 様々なクラウド基盤に構築が可能であり、エンタープライズレベルのセキュリティ要件にも対応可能です。
技術コンサルティング	生成AIを活用するうえでの技術面でのお悩みを解決します どのような技術を使うべきかの選定や実装を支援します
システム インテグレーション	生成AIを活用したWebサービスやアプリケーション、社内ツールなどの構築 を行います。 事業案の壁打ちからシステム開発、運用まで全領域でご支援が可能です。
活用を推進する教育支援	社内で活用を推進するための教育支援を行います。

生成AIの支援事例



コクヨ株式会社様

「KOKUYO DIGITAL ACADEMY」生成AI環境の構築支援

デジタル人材教育・実践プログラムとカルチャーを技術で下支え



株式会社オプテージ様

生成AIチャットボット導入で社内のお問合せをサポート

業務改善の取り組みやAI利用のナレッジ蓄積にも貢献



株式会社くすりの窓口様

RAGを活用した生成AIチャットボット提供

LLM選定やインフラを1カ月で構築



本日のゴール

1. 自社データ活用において、生成AI（LLM）を利用検討するポイントが分かる
2. 生成AI（LLM）に自社データについて答えさせるときに発生するデータマネジメント課題と対策について知識が得られる

生成AI活用における考え方のポイント | データ活用編

社内に眠る宝の山

企業が持つ**非構造化データ**の割合

90%

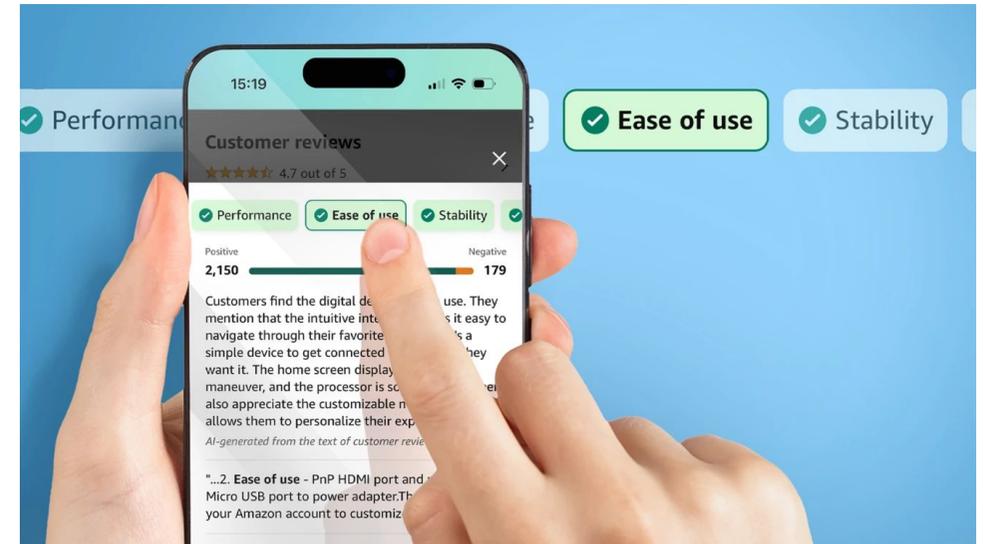
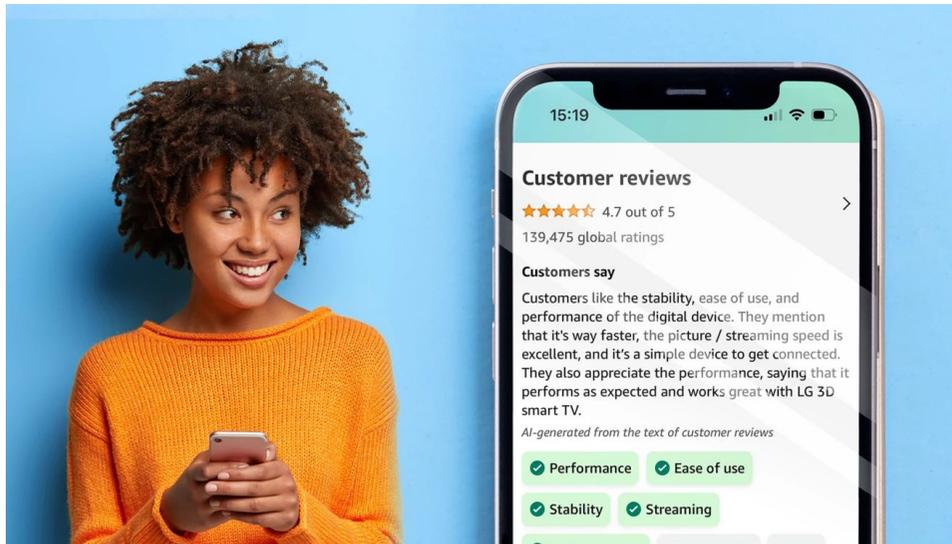
出展：blog.box.com | ”90% of your data is unstructured — and it’s full of untapped value”
<https://blog.box.com/90-your-data-unstructured-and-its-full-untapped-value>

構造化データ・非構造化データの違い

	非構造化データ	構造化データ								
概要	データ内に規則性に関する区切りがなく、データ(またはデータの一部)を見ただけで、二次元の表形式(Excelなど)に変換できないことがわかるもの	二次元の表形式(Excelなど)になっているか、データの一部を見ただけで、二次元の表形式への変換が可能であることや、変換方法がわかるもの								
形式	規則性に関する区切りのないテキスト、PDF、音声、画像、動画など	CSV、Excel など								
例		<table border="1"><thead><tr><th>世帯名</th><th>大人1</th><th>大人2</th><th>子供1</th></tr></thead><tbody><tr><td>山田家</td><td>世帯主</td><td>妻</td><td>長女</td></tr></tbody></table>	世帯名	大人1	大人2	子供1	山田家	世帯主	妻	長女
世帯名	大人1	大人2	子供1							
山田家	世帯主	妻	長女							

引用：tableau.com | ”データレイクとは？活用で得られるメリットや有用な業種などを解説”
<https://www.tableau.com/ja-jp/learn/articles/data-lake>

Amazonにおける活用例



- 信頼できるユーザーのレビューに基づきAIが製品レビューのハイライトを生成
- 製品の特徴を示すアイコンをAIが生成（性能、使い易さ、安定性など）

- アイコンを選択すると、関連するレビューテキストの詳細をまとめて読める
- 出品者の作業負担を軽減するAIのサービスも開始し、商品の効率的な販売を支援

出展：Amazon.com | “How Amazon continues to improve the customer reviews experience with generative AI”
<https://www.aboutamazon.com/news/amazon-ai/amazon-improves-customer-reviews-with-generative-ai>

ビジネスへのインパクトは多大

企業の競争優位性確保につながる情報が90%の非構造化データに眠っている

メリット

具体例

プロセス自動化

- 手作業で行っている業務を自動化して効率や生産性が上がる。
- メール分類作業、請求書処理の自動化、契約文書のレビューなどアナログに行っているビジネスプロセスを自動化できる。

意思決定の質向上

- 非構造化データの解析によって、より良い意思決定ができるようになる。
- ニュースや市場レポートの解析して投資リスクや市場トレンドを評価する。

顧客理解の深化

- 顧客の行動や感情を深く理解することによる、カスタマーエクスペリエンスの向上につながる。
- 顧客の使用写真を分析し、実際の使用方法や期待される機能を理解することで、新製品開発や既存製品の改良に役立てる。

なぜ生成AI・大規模言語モデルなのか

理由その1

言葉同士の距離関係（類似度）を学習
して近しい遠いが分かることで、非構
造化データをグルーピング、ソート、
フィルタリング、検索することが容易に
なる

理由その2

膨大な量を学習しているLLMはありとあ
らゆる次元（規則性の軸）において類似
度を学習しており、従来の機械学習や
ディープラーニングと比べて汎用性が高
い

ポイント！

莫大な労力や手間がかかる非構造化データの取り扱いが、
生成AI・LLMの登場によって低コストで実現できるようになる

生成AIに自社データを回答させるときの 課題と対策

従来の生成AIの制限とRAGの強み

RAGの活用により生成AIは事前学習した知識のみならず、追加のデータを利用して回答が可能となった

従来の生成AI (LLM) の特徴

- 回答可能な情報は学習時点までの汎用的な情報が対象
- 社内情報など一般公開されていない情報については回答ができない
- 自社データを利用できないため、業務活用は限定的になりがち

RAGの特徴

- 自社のデータや情報をLLMを組み合わせることで活用を行う
- 追加で情報を付与すれば、ドキュメントを参照しファクトに基づいた回答が可能
- 自社独自の生成AIとして、様々な業務に活用できる

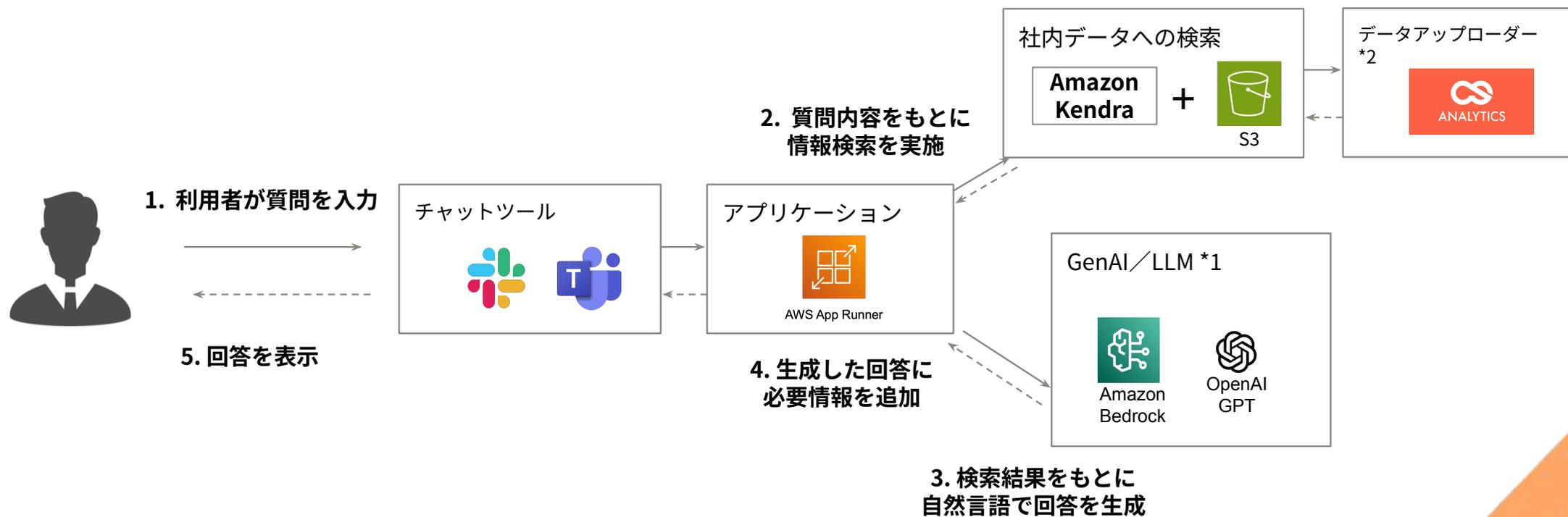
RAGとはLLMに追加で情報を与えて回答を行わせる手法



RAGアプリのデモ

システム構成図

AWS環境にて構築を実施し貴社のご要望に応じた
ユーザーインターフェースとLLMの選択が可能



*1 Azure OpenAIの利用も可能

*2 データアップローダーは CS Analytics 無料版を利用

他手法の比較（独自情報をLLMに答えさせる方法）

	構築の 難易度	構築 コスト	精度 (理想)	分析・改善の 難易度 (≒ 開発コスト)	ランニング コスト
(簡単な) fine-tuning	○ 低い	○ 安い	○	× 難しい (コスト高い)	○ 安い
RAG	○ 低い	○ 安い	○	○ やりやすい (コスト低い)	△ (検索サービスが 必要)
モデルの独自 カスタマイズ (継続学習)	× 高い	× かなり高い	◎	× 難しい (コスト高い)	○ 安い

※ 大まかな傾向の比較です。ケースや基準によって異なります

RAGの注意点（勘違いされやすい点）

全部のドキュメントを”学習”するわけではない

- ・ 検索でヒットした一部の文章のみをもとに回答する

なので、回答の質は

ドキュメントを初見の人が理解できる範囲で答える感じ

= 新入社員の人が回答する感覚

RAG利用時に発生する課題と解決の例

課題例1: CSVファイルの途中が抽出されてしまった

・解決策: ファイルの分割

課題例2: 似たようなドキュメントの内容が混同して回答された

・解決策: タグ付け・フィルタリング

課題例3: パワポファイルの読まれ方が意図しない形になってしまった

・解決策: マルチモーダルなモデルを使って読み込む

課題例4: ドキュメントに書かれてない社内知識が必要だった

・解決策: 用語集をつくる

課題例 1：対応しない箇所をもとに回答をしてしまった

タイトル	回答例	詳細情報・回答根拠・リンク先情報
ログインパスワードを忘れたときの手順を教えてください	ログインパスワードが分からなくなった場合は Slackで情シス宛に初期化依頼をしてください。 詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/1)	初期化する権限は情シスのみが保有しています。ユーザのみでは行えないので、ご連絡ください。
〇〇が反応しないときの対処手順を教えてください	設定画面を開き、〇〇が有効になっているか確認してください。確認方法や詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/2)	〇〇は誤って無効にされてしまうケースが多いです。 設定を改めてご確認ください
〇〇できないときの対処方法を教えてください	ソフトウェアの再起動をしてください。詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/3)	再起動処理が完了するまでに、数分かかります。再起動後アイコン上はすぐに接続できるように見えますが、 処理中の状態ですので、時間をあけて接続して下さい
...

質問

「〇〇できないときの対処方法を教えてください」



回答

「設定画面を開き、有効になっているか確認 ...」

課題例 1 の原因：抽出する範囲が合っていない

タイトル	回答例	詳細情報・回答根拠・リンク先情報
ログインパスワードを忘れたときの手順を教えてください	ログインパスワードが分からなくなった場合は Slack で情シス宛に初期化依頼をしてください。 詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/1)	初期化する権限は情シスのみが保有しています。ユーザのみでは行えないので、ご連絡ください。
〇〇が反応しないときの対処手順を教えてください	設定画面を開き、〇〇が有効になっているか確認してください。確認方法や詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/2)	〇〇は誤って無効にされてしまうケースが多いです。設定を改めてご確認ください
〇〇できないときの対処方法を教えてください	ソフトウェアの再起動をしてください。詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/3)	再起動処理が完了するまでに、数分かかります。再起動後アイコン上はすぐに接続できるように見えますが、処理中の状態ですので、時間をあけて接続して下さい
...



タイトル 回答例 詳細情報・回答根拠・リンク先情報
ログインパスワードを忘れたときの手順を教えてください
"ログインパスワードが分からなくなった場合は Slack で情シス宛に初期化依頼をしてください。詳細はリンク先を参照してください。
[リンク先](http://sample.classmethod.jp/sample/sample/1)"
"初期化する権限は情シスのみが保有しています。ユーザのみでは行えないので、ご連絡ください。"
〇〇が反応しないときの対処手順を教えてください
"設定画面を開き、〇〇が有効になっているか確認してください。確認方法や詳細はリンク先を参照してください。
[リンク先](http://sample.classmethod.jp/sample/sample/2)"
"〇〇は誤って無効にされてしまうケースが多いです。設定を改めてご確認ください"
〇〇できないときの対処方法を教えてください
"ソフトウェアの再起動をしてください。詳細はリンク先を参照してください。
[リンク先](http://sample.classmethod.jp/sample/sample/3)"
"再起動処理が完了するまでに、数分かかります。再起動後アイコン上はすぐに接続できるように見えますが、処理中の状態ですので、時間をあけて接続して下さい"

行をまたいで、途中が抽出されしまった
ヘッダーの情報もない

課題例 1 の解決方法：ファイルを小さく分割する

タイトル	回答例	詳細情報・回答根拠・リンク先情報
ログインパスワードを忘れたときの手順を教えてください	ログインパスワードが分からなくなった場合は Slackで情シス宛に初期化依頼をしてください。 詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/1)	初期化する権限は情シスのみが保有しています。ユーザーのみでは行えないので、ご連絡ください。
〇〇が反応しないときの対処手順を教えてください	設定画面を開き、〇〇が有効になっているか確認してください。確認方法や詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/2)	〇〇は誤って無効にされてしまうケースが多いです。設定を改めてご確認ください
〇〇できないときの対処方法を教えてください	ソフトウェアの再起動をしてください。詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/3)	再起動処理が完了するまでに、数分かかります。再起動後アイコン上はすぐに接続できるように見えますが、処理中の状態ですので、時間をあけて接続して下さい
...

FAQ.csv

タイトル	回答例	詳細情報・回答根拠・リンク先情報
ログインパスワードを忘れたときの手順を教えてください	ログインパスワードが分からなくなった場合は Slackで情シス宛に初期化依頼をしてください。 詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/1)	初期化する権限は情シスのみが保有しています。ユーザーのみでは行えないので、ご連絡ください。

FAQ_0.csv

タイトル	回答例	詳細情報・回答根拠・リンク先情報
〇〇が反応しないときの対処手順を教えてください	設定画面を開き、〇〇が有効になっているか確認してください。確認方法や詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/2)	〇〇は誤って無効にされてしまうケースが多いです。設定を改めてご確認ください

FAQ_1.csv

タイトル	回答例	詳細情報・回答根拠・リンク先情報
〇〇できないときの対処方法を教えてください	ソフトウェアの再起動をしてください。詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/3)	再起動処理が完了するまでに、数分かかります。再起動後アイコン上はすぐに接続できるように見えますが、処理中の状態ですので、時間をあけて接続して下さい

FAQ_2.csv

課題例 1 の結果：正しく読み込まれた

タイトル	回答例	詳細情報・回答根拠・リンク先情報
〇〇できないときの対処方法を教えて	ソフトウェアの再起動をしてください。 詳細はリンク先を参照してください。 [リンク先](http://sample.classmethod.jp/sample/sample/3)	再起動処理が完了するまでに、数分かかります。 再起動後アイコン上はすぐに接続できるように見えますが、 処理中の状態ですので、時間をあけて接続して下さい

質問

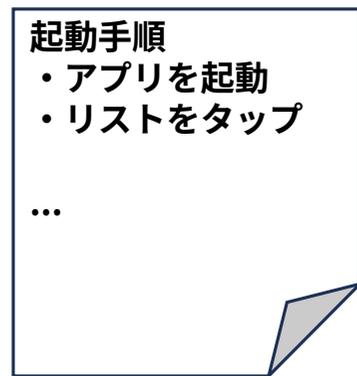
「〇〇できないときの対処方法を教えて」



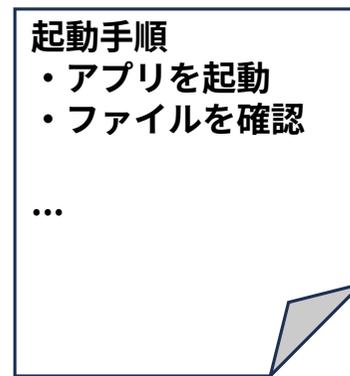
回答

「ソフトウェアの再起動をしてください ...」

課題例 2 : 似たドキュメントを内容を混同してしまった



製品A マニュアル



製品B マニュアル

質問

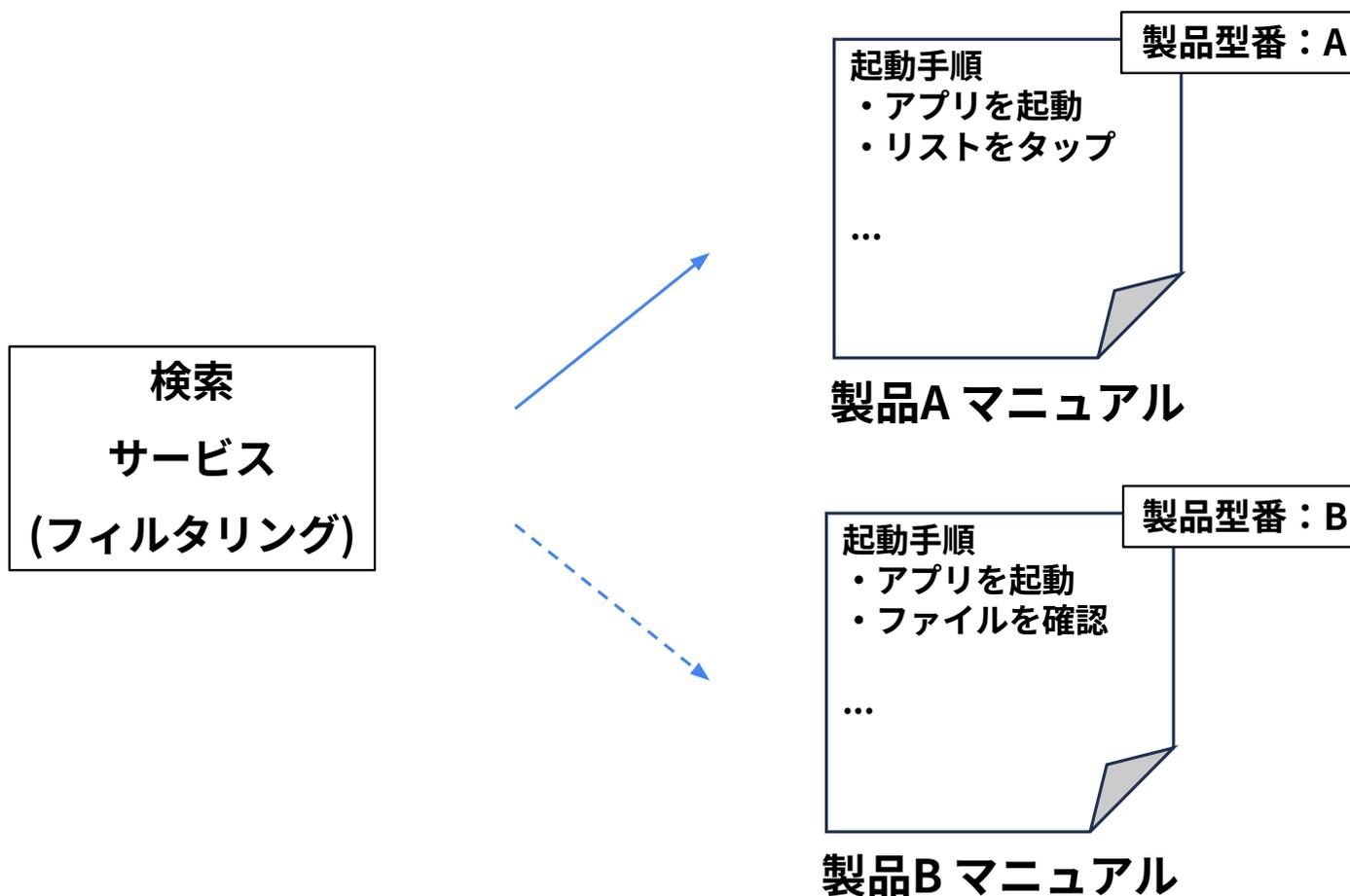
「製品Aの起動手順を教えてください」



回答

「アプリを起動し、リストをタップした後、
ファイルを確認してください」

課題例 2 の解決方法：タグ付け・フィルタリングをする



課題例 3 : 画像が読まれない

活気ある職場・働きやすい環境①

○職場風景



(資料) 採用パンフレットより抜粋

○働きやすい職場環境

- ✓ テレワーク ※29FYは延べ7,000人以上が実施。中央省庁では最多。
- ✓ ペーパーレス ※4年で37%削減
- ✓ フレックス
- ✓ 風通しのよい職場 (職員意識調査: 職場満足度7割以上)
- ✓ 様々な研修制度 (年間100回以上の勉強会の開催など)



個人PC: 軽量で持ち運びが容易²

https://www.jinji.go.jp/saiyo/siken/senkou/setsumeikai_17.pptx

※ Kendraのリファレンスにも
デフォルトでは画像が読み込まれないことは明記されています

setsumeikai_17

日本経済・国民の暮らしを豊かにする1活気ある職場・働きやすい環境① ○職場風景 (資料) 採用パンフレットより抜粋 ○働きやすい職場環境 テレワーク ※29FYは延べ7,000人以上が実施。中央省庁では最多。ペーパーレス ※4年で37%削減 フレックス 風通しのよい職場 (職員意識調査: 職場満足度7割以上) 様々な研修制度 (年間100回以上の勉強会の開催など) 個人PC: 軽量で持ち運びが容易

https://s3.ap-northeast-1.amazonaws.com/generat.../setsumeikai_17.pptx

▶ Document fields

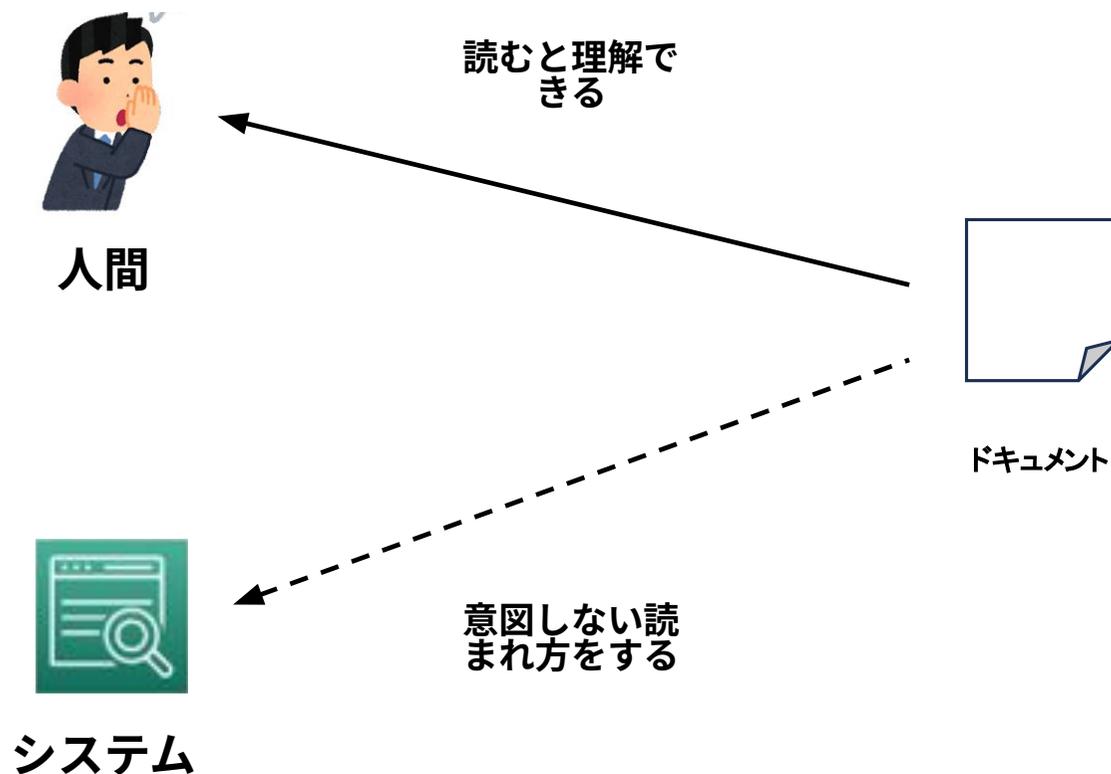


そもそも画像があったかどうかも
わからない

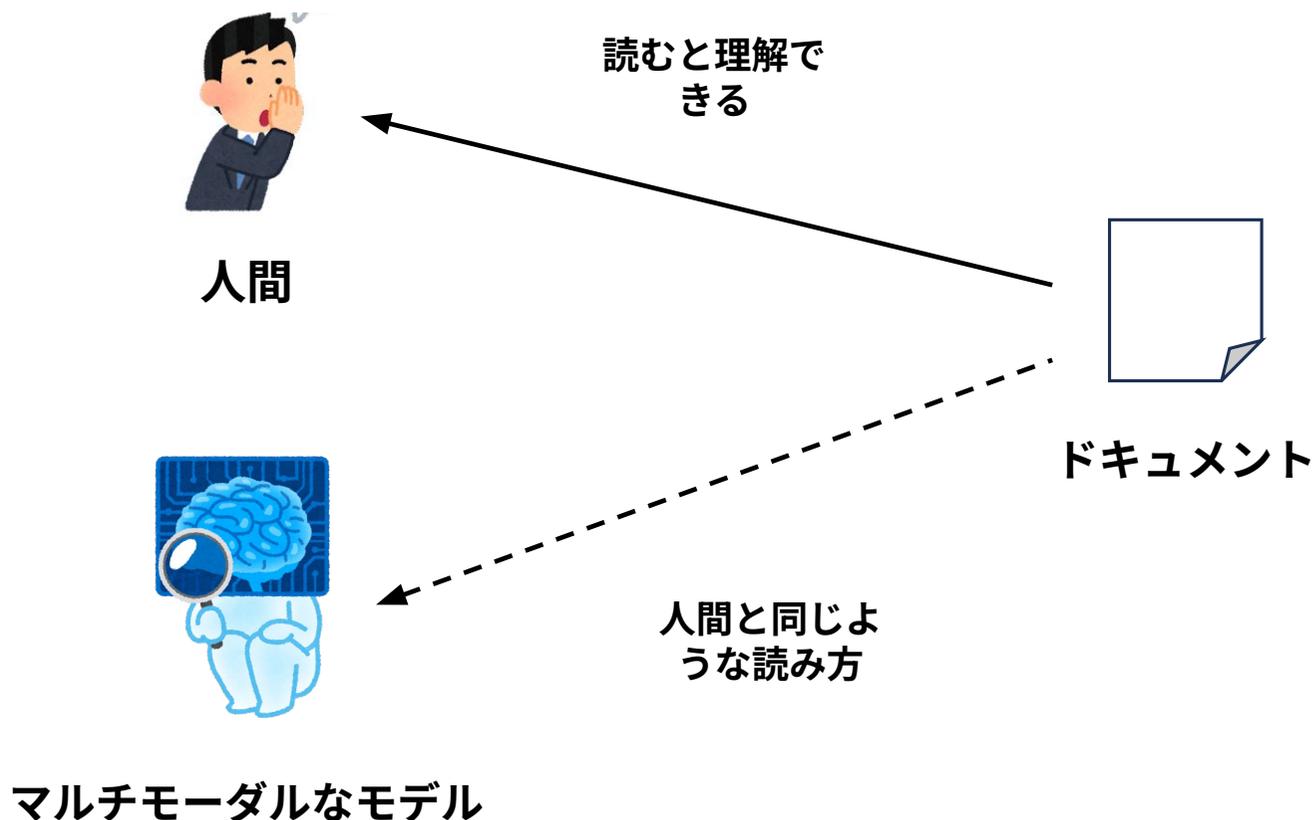
補足: 既存のドキュメントローダーでは、
画像は読み込まれるものの、
変換後のテキストは簡素で情報が足りないことが多い

課題例 3 の原因：人間の読み方とシステムの読み方が異なる

人間が読んでわかりやすい \neq システムが読み込んだあとの形式がわかりやすい



課題例 3 の解決方法：マルチモーダルなモデルを使う



課題例3の結果：人間が読む順序で文字起こしできた



※ Claude 3



経済産業省のMission

日本経済・国民の暮らしを豊かにする

産業政策

- 人工知能、IoT、ヘルスケア
- データ活用、中小企業
- 産業構造・・・

通商・貿易

- EPA、TPP、インフラ輸出
- 新興国戦略、ルール形成
- 戦略・・・

資源・エネルギー

- 電力自由化、新エネ・省エネ
- 原発、資源外交・・・

手段

- 経済成長
- 産業競争力の強化
- イノベーション
- 世界の富の取り込み
- エネルギー安定供給

目的

- 社会課題の解決
Ex. 少子高齢化、貧困問題、
世界の不安定化
- 豊かな社会の実現

詳細はこちらのブログをご覧ください

<https://dev.classmethod.jp/articles/read-powerpoint-document-with-claude-3/>

課題例3の結果：画像を説明させることができた

活気ある職場・働きやすい環境①

○職場風景



(資料) 採用パンフレットより抜粋

○働きやすい職場環境

- ✓ テレワーク ※29FYは延べ7,000人以上が実施。中央省庁では最多。
- ✓ ペーパーレス ※4年で37%削減
- ✓ フレックス
- ✓ 風通しのよい職場 (職員意識調査：職場満足度7割以上)
- ✓ 様々な研修制度 (年間100回以上の勉強会の開催など)



個人PC：軽量で持ち運びが容易²

✳ Claude 3



活気ある職場・働きやすい環境 1

職場風景

[オフィスの様子が写っている。複数の人が机を囲んで作業をしている。]

[3人の男女がパンフレットを見ながら話し合っている。壁には絵画が掛かっている。]

働きやすい職場環境

- テレワーク ※29FYは延べ7,000人以上が実施。中央省庁では最多。

- ペーパーレス ※4年で37%削減

- フレックス

- 風通しのよい職場 (職員意識調査：職場満足度77割以上)

- 様々な研修制度 (年間100回以上の勉強会の開催など)

[2台のノートPCが写っている。]

個人PC:軽量で持ち運びが容易

※ プロンプトの指示は簡易なものを使用したため、出力される説明は改良の余地があります

詳細はこちらのブログをご覧ください

<https://dev.classmethod.jp/articles/read-powerpoint-document-with-claude-3/>

課題例 4 : 社内知識・業務知識が必要

例 :

- ・ 質問

「**20期**の年末年始の
スケジュールを教えて」

- ・ ドキュメント

- ・ 2023年の年末年始

- ・ 2022年の年末年始

ポイント

- ・ 20期が何なのか把握させる

- ・ 20期が何年に対応するのか
計算させる

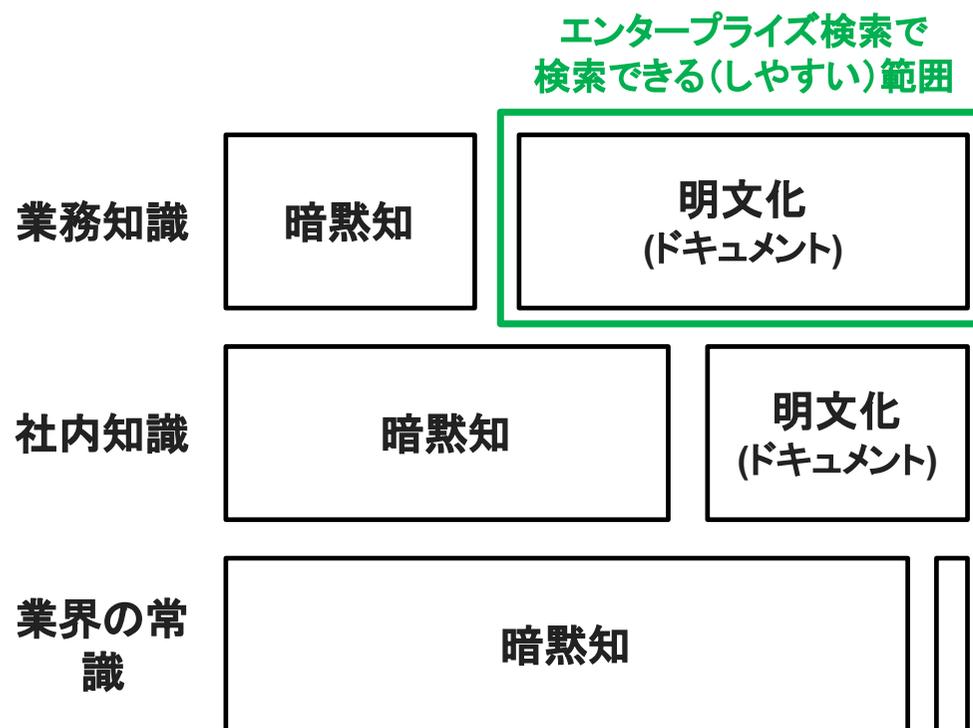
- ・ 1期が何年なのか教える

普遍的な社内知識に対応させる

(こうしたケースが大量にある)

課題例 4 の原因：情報量の差

回答に関する情報(社内情報に関するQAの場合)

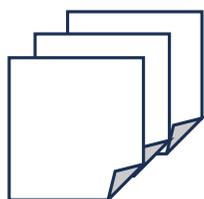


システムが使用している情報は、人間に比べてごく一部

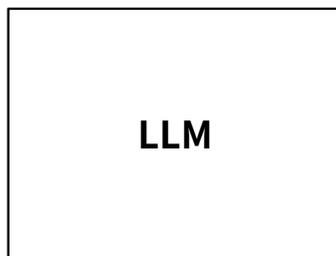
→ 暗黙的な情報が隠れている

課題例 4 の解決策(の 1 つ)：用語集をつくる

[# ドキュメント]のテキストの中から、社内用語集を作成したいです。
社内用語を抽出し、その単語の意味も推測して出力してください。
出力は[# 出力の形式]に従ってください。



ドキュメント



20期：2023年7月～
2024年6月のこと

用語と意味の
リスト

課題例 4 の結果：社内用語とその説明を抽出できた

2255ファイルから5153件を抽出できた（※ 重複あり）

一例

#01help-me	社内のヘルプデスクやサポート窓口を指す用語。社員がIT関連のトラブルや問題を報告するための連絡先
第14期	会社の年度を表す用語。会社の創業年を1期として、その年から数えて14年目を表す。2017年度を指す。
入場アンケートフォーム	新入社員や異動者が入社・異動時に提出するフォーム。必要なPCやシステムのアクセス権限などを申請するために使用される。
Nulabアカウント	Cacooを利用するために必要なアカウント。社内の手続きを経て発行される
2代目めそ子	社内用語で、社内向けイベントカレンダーにイベントを追加する際に招待するゲストの名前。社内システムやツールを指している可能性がある。
04.パソコン・周辺機器利用・返却申請WF	社内のパソコンや周辺機器の利用申請や返却手続きを行うためのワークフロー。
Lab	チームでワイワイ開発したり、もくもくと作業をするエリアの名称
Hub	ビジネスについて議論するエリアの名称。電話やオンライン会議も可能
Core	財務経理室、労務の執務室、倉庫、セキュリティルームのエリアの総称
キャリアブレーション	評価の基準を合わせるための会議。評価者間で評価のばらつきを防ぐために行われる
JD	社内の人事評価制度に関連する用語。職務記述書（Job Description）の頭文字から取った略称で、目標設定や評価の面談やシートを指す。
FU-1102(どんたく)	福岡オフィスの11階にある会議室の名称。4人収容可能

RAG環境構築サービスのご紹介

生成AIチャットボット 「らくらくRAG導入パック」

AIチャットボットで問い合わせ対応削減！

問い合わせ対応のこんなお悩みありませんか？

問い合わせ対応に
時間がかかり、
担当者の業務負荷
が高い

社内の情報が散在
しており、
適切な回答を素早
く見つけるのが
難しい

回答品質にばらつ
きがあり、利用者
の満足度が低い

その悩み、「らくらくRAG導入パック」
で解決できます！

解決1

チャットボットによる自動応答で問い合わせ対応時間を大幅に短縮

解決2

社内ドキュメントを一元管理し、生成AIが自然言語で回答を生成

解決3

AIによる高品質な回答で利用者満足度の向上を実現



RAG導入支援事例:くすりの窓口様

- 技術力とスピード感を大事にしたいという考えからパートナーをクラスメソッドに決定
- 約1カ月の開発期間を経てRAGを活用した生成AIチャットボットのPoC検証用の環境を構築
- 社内チャットで質問するとそのサービスについて回答が得られる仕組み
- クライアントに出せるほどのクオリティになれば、問合せ用のシステムに組み込む予定

サービスイメージ



最短二週間で社内問い合わせを効率化するRAG環境を構築！

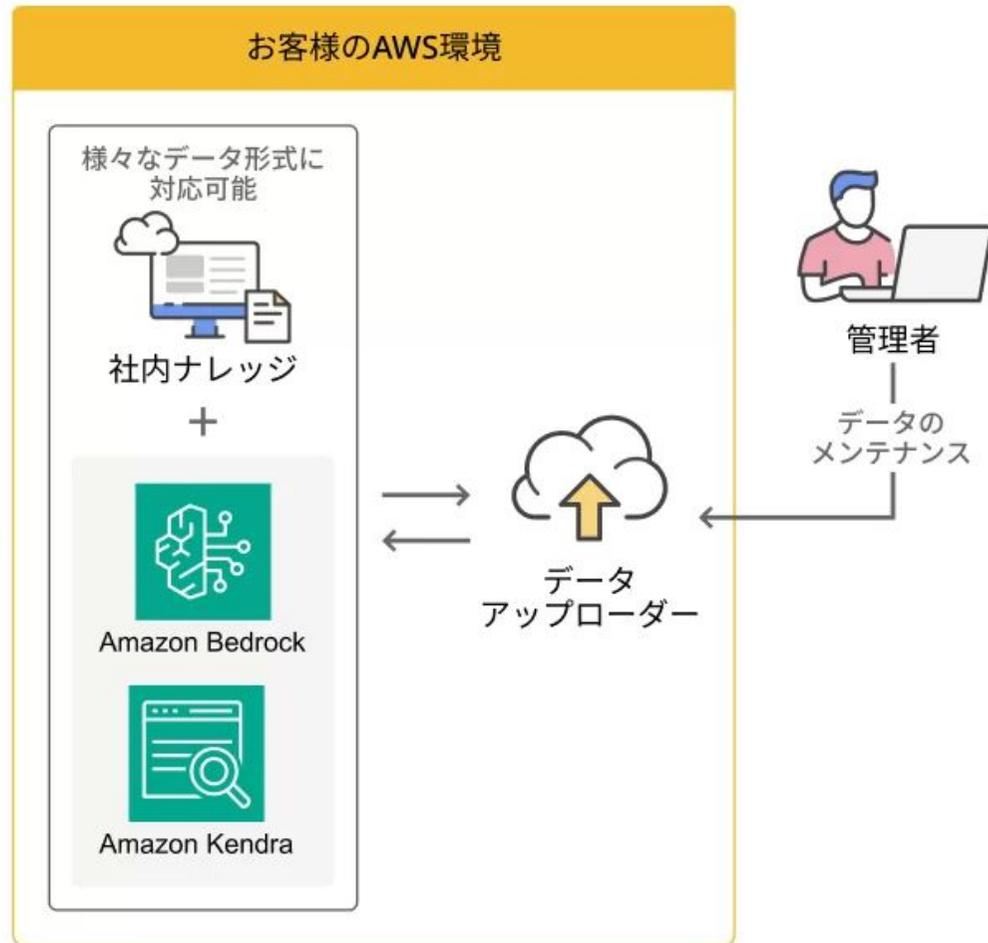
有給休暇申請について教えてください



ユーザー

Microsoft Teams
slack

有給休暇申請の方法は...



らくらくRAG導入パックの特徴



最短2週間で構築

事前に社内データを準備している場合、最短2週間で導入が可能



社内チャットツールと連携して利用が可能

Slack, Teamsなどのチャットツールと連携して利用できる



様々なLLM (大規模言語モデル) と連携

AWS Bedrock, OpenAI, Azure OpenAI など様々なモデルで構築が可能



自社でデータが管理できる セキュアな環境

自社のクラウド環境で構築するため、安心して利用可能



手軽にデータ管理

専用のアップローダーを利用することで、簡単にデータ追加や削除が可能です。



回答内容とフィードバックの 確認が可能

利用～分析～改善のPDCAサイクルを回すことができます。

期待される業務効果

クラスメソッドでは生成AIを活用したチャットボットを導入し、社員の情報検索や問い合わせ業務の効率化を実現

2023年10月03日 プレスリリース

**クラスメソッド、生成AIと社内データの連携により70%
超のヘルプデスク回答満足度を実現**
～Amazon Bedrock、Amazon Kendraなどを自在に組み合わせて社内業務をアシストするAI環境を構築～

この記事シェア   

クラスメソッド株式会社（本社：東京都港区、代表取締役社長：横田 聡、以下「クラスメソッド」）は、社内問い合わせ業務に生成AIを利用したチャットボットを構築・導入し、総務・労務・経理・法務・情報システムなど多岐にわたる社内問い合わせに対して70%を超える満足度を実現しました。

このたび構築したチャットボットは、ユーザーからの質問に対して、社内に点在する複数のドキュメントをふまえて回答を作成するものです。AIの精度を高めるための教師データの作成などの構築コストはほとんどかけておらず、また、既存ドキュメントを活用することでQAの情報アップデートなどの運用についてもユーザーに負担をかけることなく対応することが可能です。

社内情報を活用することにより、導入・運用コストを抑え、短期間かつ安価に、問い合わせ窓口チャットボットを導入することができました。

まとめ

- 生成AI・LLMの登場により非構造化データを活用する敷居が下がった
- 生成AIの活用を検討するときは、非構造化データを加工・変換することで高い効果が得られる業務課題を見つける
- RAGのアーキテクチャでは、検索される前にデータの前処理を加えると生成される回答の品質が上がる

The background is a solid orange color with a subtle, light-colored diamond-shaped grid pattern. Scattered throughout the background are various 3D geometric shapes in shades of orange and brown, including a cylinder, a sphere, a ring, and several rectangular prisms of different sizes and orientations.

Classmethod Showcase